Implementation of an Enhanced VoIP Codec Transcoder to Improve VoIP Quality in IP Telephone Infrastructure

Thubelihle S. Zulu, Topside E. Mathonsi

***Abstract*— Poor Voice over Internet Protocol (VoIP) quality in IP telephone infrastructure is a major concern and it can affect business growth, especially for businesses that deal with interacting with a client over the phone. Speech or audio signals are usually affected by codec mismatch which leads to packet loss and jitter. VoIP telephone system is growing at a rapid speed and has received much attention because of its call cost internationally and nationally and fewer resources needed compared to traditional voice telephone systems or Public Switched Telephone Networks (PSTN). The main aim of this paper was to develop enhanced voice quality in VoIP platform systems. The proposed solution was achieved by implementing the amended VoIP codec transcoding system that auto negotiates VoIP codec with the intention of preventing VoIP codec mismatch via standalone and software VoIP codec transcoding systems. Session Initiation Protocol (SIP) phone, Private Branch Exchange (PBX) systems, and Session Border Control (SBC) were used in order to implement and evaluate the proposed solution. The simulation results have shown that the proposed solution has reduced packet loss and jitter thereby improving voice quality.**

Index—Session Border Control, Private Branch Exchange, Public Switched Telephone Network, Voice over Internet Protocol, and Session Initiation Protocol.

I. Introduction

VoIP codec is the primary rule to the functioning of the Voice over Internet Protocol (VoIP) and Public Switched Telephone Network (PSTN). The main goal of VoIP codec is to replace audio signals in digital form with a few bits while maintaining the accessibility and voice quality essential for the specific application. VoIP telephone system is growing at a faster speed and has received much attention because of its cost saving of about 50% as compared to traditional voice telephone systems [1-3].

Speech signals are usually affected by codec mismatch, packet loss, and jitter which affect the voice quality. Voice services are based on different internet protocols thus, voice codecs mismatch.

Manuscript received January 10, 2023; revised April 8, 2023.
This paper is funded by the Tshwane University of Technology. Thubelihle S. Zulu is a student at Tshwane University of Technology, South Africa (email: sackyzulu@gmail.com).
Topside E. Mathonsi is a senior lecturer at Tshwane University of Technology, South Africa (email: MathonsiTE@tut.ac.za).

VoIP platforms implemented in the wireless network are not the best option for voice quality. This is because wireless is not stable as wired network infrastructure thus, it is not reliable [4-6]. VoIP telephone systems and hosted VoIP telephones are easy to install and maintain thus, are easy to integrate with new business applications. However, upgrading from traditional telephone systems to VoIP platform are facing the inherent voice quality, packet loss, codec mismatch, and calls dropping issues of IP telephone systems.

Due to technological changes, the traditional telephone system requires a lot of money in terms of call cost and resources compare to the VoIP telephone system which normally uses a session border gateway [7-12]. This paper introduced the amended transcoding of VoIP codecs in order to enhance voice quality and prevent VoIP codec mismatch. Hence, the proposed solution reduced packet loss, packet delay, jitter, and voice echo.

II. Proposed Solution

This paper proposed the VoIP codecs transcoding system for both the receiver and the sender. This was done in order to ensure that codec mismatch was minimized while improving the VoIP quality.

A. *Figures VoIP Architecture*

VoIP is a mode of communication in which a user can make phone calls over a broadband internet connection rather than traditional analog phone lines. VoIP network has five major components as explained below and in Figure 1 which is the general VoIP system architecture.

- VoIP phones, softphones, PC applications, and other devices from which end users initiate and receive VoIP calls.
- PABX which manages all VoIP control connections.
- PSTN VoIP gateways, which convert voice content for transport over the IP network.
- IP Network, which transports the audio payload and video data streams.
- SBC controls real-time, session-based traffic at the signaling or calls control and transport layers as it crosses network borders and network domains.

The below structure and steps represent where and how the transcoding takes place, the caller initiates the call with the codec G.711 which. Which is normally used or mostly

ISBN: 978-988-14049-4-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

prioritized in the VoIP Platform due to a better quality of speech.

- Initialization: Initialize the call with one codec and other alternative codecs
- Transcoder System: Acknowledge the VoIP codec
- Call Signaling: Establish a call signaling connection to call using a codec and called check the codec whether is supported or not
- Negotiation: Negotiate the codec with the transcoder system
- Transcoder System: acknowledge the codec and start the RTP media or renegotiate the codec
- End: Terminate the call

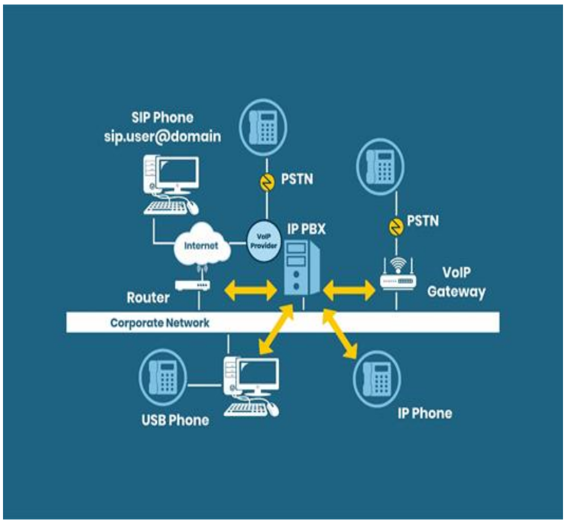

Figure 1. General VoIP system architecture

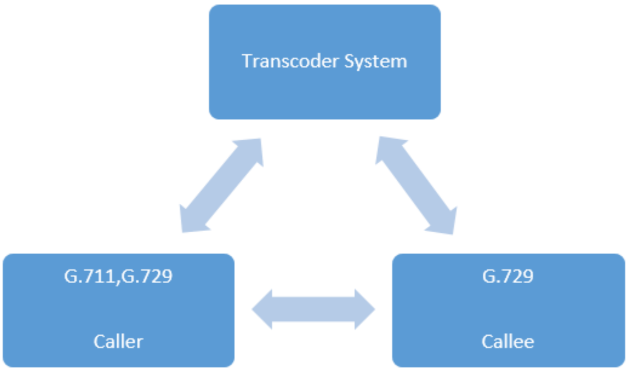

Figure 2. VoIP transcoding system structure

B. *Electronic Image Files*

This study selected and enhanced the mostly used transcoder algorithm which is the dynamic codec selection transcoder algorithm. This is because the algorithm measures the delay and packet loss in order to determine the better voice quality and choose the correct codec. Hence, VoIP systems are mostly affected by delay and packet loss, and codec mismatch during the call.

ITU-T E model represents an analytical model of voice quality defined in the ITU-T recommendation G.107.E model provides a framework for real-time online quality estimation from network performance measurement. The $\boldsymbol{R}$ factors are calculated as follows and translated into $\boldsymbol{MOS}$ scale through the following expressions. The $\boldsymbol{R}$ factor is defined as in equation (1) :

$$\boldsymbol{R} = \boldsymbol{R_0} - \boldsymbol{I_s} - \boldsymbol{I_d} - \boldsymbol{A} \qquad (1)$$

$\boldsymbol{R_0}$ represent the basic signal-to-noise ratio, $\boldsymbol{I_s}$ reflects the impairments occurring simultaneously with the voice signal due to quantization. It is a function of several parameters, none of which are related to the underlying packet transport. $\boldsymbol{I_s}$ can be defined as in equation (2):

$$\boldsymbol{I_s} = \boldsymbol{I_e} + (\mathbf{95} - \boldsymbol{I_e}) * \frac{\boldsymbol{P_{pl}}}{\boldsymbol{P_{pl}} + \boldsymbol{B_{pl}}} \qquad (2)$$

Where $\boldsymbol{P_{PI}}$ is the packet loss ratio and is a codec-dependent parameter related to the capacity of supporting not randomly distributed packet loss. $\boldsymbol{I_e}$ represent the parameters for the main ITU-T codecs. Estimating the impact of end-to-end delay. $\boldsymbol{D\_codec}$ represent codec-related delays caused by packetization and encoding, including processing ad lookahead. To determine the $\boldsymbol{D\_codec}$ it is necessary to calculate the packetization delay which is defined as in equation (3):

$$\boldsymbol{D_{codec}} = \frac{\boldsymbol{frames}}{\boldsymbol{packet}} * \boldsymbol{framesize} + \boldsymbol{lookahead} + \boldsymbol{process} \qquad (3)$$

Aside from selecting codec based on bit rate, it is also important to note algorithmic delays such as lookahead and processing complexity. $\boldsymbol{I_d}$ is expressed as equation (4) and further equation (5), where $\boldsymbol{d}$ represents the one-way delay (in milliseconds). Voice quality degrades more rapidly when the delays exceed 177.3 ms. $\boldsymbol{I_d}$ can be defined as in equation (4):

$$\boldsymbol{I_d} = (\mathbf{0.024}\boldsymbol{d}) + \mathbf{0.11}(\boldsymbol{d} - \mathbf{177.3}).\boldsymbol{I}(\boldsymbol{d} - \mathbf{177.3}) \qquad (4)$$

$$\boldsymbol{I} = \begin{cases} \mathbf{0}, & x < 0 \\ \mathbf{1}, & otherwise \end{cases} \qquad (5)$$

Once the R factor value has been obtained, it is desired to translate it into a relative MOS score. The R factor is related to MOS in the following equations :

$$\boldsymbol{Mos} = \begin{cases} \mathbf{1} \\ \mathbf{1} + \mathbf{0.035}\boldsymbol{R} + \mathbf{7.10}^{-\mathbf{6}} * \boldsymbol{R} * (\boldsymbol{R} - \mathbf{60}) * (\mathbf{100} - \boldsymbol{R}) \\ \mathbf{4.5} \end{cases} \qquad (6)$$

Factors: $R < 0$; $0 < R < 100$; $R > 100$

ISBN: 978-988-14049-4-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Below is the dynamic codec selection algorithm:

Algorithm 1: Dynamic codec selection

Initialization:

1. Enter Min_Bitrate (M_b), Max_bitrate (M_{xb}), Max_Pl (M_{xpl}), Min Pl(M_{pl}), Codec_bitrate (i)

Process:

2. For each entry (i) in codecs do:
3. If (codec_bitrate (i) >M_{xb})
4. $R = R_0 - I_s - I_d + A$
5. delete entry,
6. endif
7. end for
8. For each entry (i) in codec do:
9. If codec_bitate (i)= estimate_bitrate (codec(i), packet loss, delay)
10. $I_s = I_e + (95 - I_e) * \frac{P_{pl}}{P_{pl}+B_{pl}}$
11. $I_d = (0.024d) + 0.11(d - 177.3).I(d - 177.3)$
12. Where I = $\begin{cases} 0, & x < 0 \\ 1, & otherwise \end{cases}$
13. endfor
14. if (codec_bitrate (i) >M_b)
15. $Mos = \begin{cases} 1 \\ 1 + 0.035R + 7.10^{-6} * R * (R - 60) * (100 - R) \\ 4.5 \end{cases}$
16. output = codec
17. endif
18. endfor

C. Electronic Image Files

This paper proposed an enhanced adaptive codec selection intending to enhance voice quality and avoid codec mismatch based on the delay value of the packet, packet size, and packet loss. Codec selection will be achieved by using the enhanced version algorithm of dynamic codec selection.

The adaptive codec selection will transcode the codec with the intention of avoiding codec mismatch during the call or before the real-time media is established using equation (1), equation (2), equation (3), equation (4), equation (5) and equation (6). The study enhanced the dynamic codec selection transcoder by adding some of the steps on the algorithm with the intention of enhancing the quality of voice thereby avoiding codec mismatch. Hence, the dynamic codec selection algorithm reduced codec mismatch.

Below is the adaptive algorithm.

Algorithm 2: Adaptive codec selection

Initialization:

1. Enter Min_Bitrate (M_b), Max_Bitare (M_{xb}), Max_Pl (M_{xpl}), Min Pl(M_{pl}), Codec_bitrate (i)

Process:

2. If (codec_bitrate (i) > = M_{xb})
3. $R = R_0 - I_s - I_d + A$
4. else
5. If (codec_bitrate (i) <= M_b)
6. endif
7. end for
8. If codec_bitrate (i)= estimate_bitrate (codec(i),packet loss, delay)
9. $I_s = I_e + (95 - I_e) * \frac{P_{pl}}{P_{pl}+B_{pl}}$ (3.2)
10. $D_{codec} = \frac{frames}{packet} * framesize + lookahead + process$
11. else
12. If codec_bitrate (i)= estimate_bitrate(codec(i),packetization, codec delay)
13. $I_d = (0.024d) + 0.11(d - 177.3).I(d - 177.3)$ (3.4)
14. Where I = $\begin{cases} 0, & x < 0 \\ 1, & otherwise \end{cases}$ (3.5)
15. endif
16. endfor
17. if (codec_bitrate (i) >M_b + D_{codec})
18. $\boldsymbol{Mos} = \begin{cases} \mathbf{1} \\ \mathbf{1 + 0.035R + 7.10^{-6} * R * (R - 60) * (100 - R)} \\ \mathbf{4.5} \end{cases}$
19. output = codec
20. endif
21. endfor

III. Results and Discussions

In this paper, researchers have used a simulation tool to carry out experimental evaluations of protocols and applications under various network configurations. Simulations are used to simulate real-time communication behavior in a network context. This is because setting up real-time communication networks takes time and might be quite expensive. The experimental assessments of the suggested adaptive codec selection algorithm are consequently discussed in this section.

ISBN: 978-988-14049-4-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

The average delay, packet loss, and average throughput are the three performance metrics that were examined in this paper.
These performance measurements are significant since the review of literature research has shown that preventing codec mismatch in any network improves the quality of service (QoS). Trace files were automatically created and utilized to capture all the findings during simulations of the suggested approach. The results of the analysis, which were achieved through numerous simulations, were displayed in the form of graphs using the R programming language, and the PESQ scale was used to determine the voice quality on a scale of 0 to 5.

A. Experimental Evaluation

This paper employed TCL scripts to build the suggested simulation architecture in NS-2. This was done in order to assess the impact of introducing the adaptive codec selection algorithm in order to improve the dynamic codec selection algorithm. The NAM interface was utilized to monitor packet losses, node positions, and packet transfer during the simulations.

The simulation tool was set up and used using Ubuntu 20.04.5 LTS running on Windows 10 Home (WSL). C++ and OTcl programs were used to implement the suggested adaptive codec selection technique and the network topology. For this investigation, detailed simulations were run multiple times to produce the most promising and trustworthy results. The advantage of NS-2 is that it generates trace files on its own to record outcomes like packet transmissions among many others. This paper used the R programming language to graphically depict the outcomes from the AWK scripts.

B. Performance Metrics

The dynamic codec selection was compared to the proposed adaptive codec selection. The dynamic codec selection algorithm was selected because proposed to reduce packet loss and delay while improving the quality of voice. The adaptive codec selection technique can automatically negotiate codec with other VoIP gateways in order to prevent codec mismatch. The review of the literature indicated that assigning network nodes distinct priority levels ensures a particular level of performance during data transmissions. In addition, the dynamic codec selection process depends on bandwidth and packet loss for adequate data transmission.

The proposed adaptive codec selection algorithm's performance was assessed using the performance metrics listed below.

1. Average throughput – the measurement of how much information can be successfully transported between two or more network nodes in a given amount of time.
2. Average delay – the measurement of how long data must travel across a network from one node to another.
3. Packet loss – happens when one or more data packets traveling over a computer network are unable to reach their intended location as a result of data transmission issues or network congestion.

C. Average Throughput

The AWK program was developed and used to determine the average network throughput. In this paper, the adaptive codec selection technique used the VoIP codecs G.711 and G.729, which produced a promising average network throughput (See Figure 3). The adaptive codec selection algorithm determines prioritization limitations for all relay nodes and assigns each node a certain priority level based on workflow operations, is the cause of this. On the other hand, the proposed method in this paper allows a maximum number of customers to access the network's resources and services in priority order, which improves QoS.

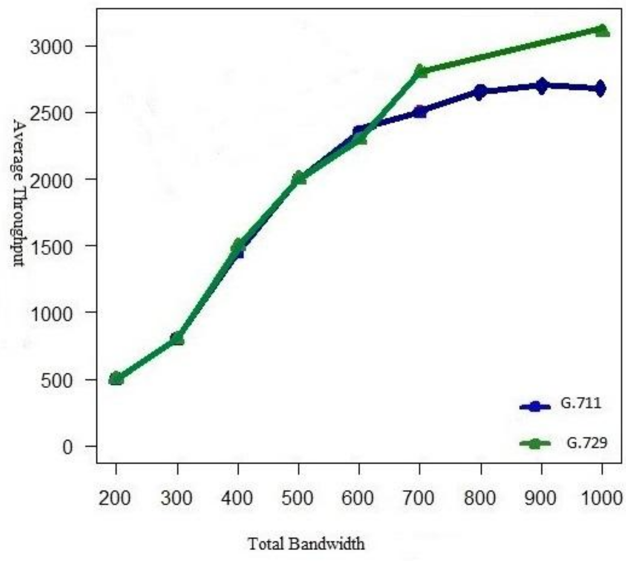

Figure 3. Average throughput

D. Average Delay

This paper compares the most popular VoIP codecs and uses the adaptive codec method to determine the average end-to-end delay. As presented in Figure 4, the adaptive algorithm reduced the average end-to-end delay. The average delay or end-to-end delay was made possible by the network's minimal congestion, which is achieved by allowing the greatest number of clients to connect to it in a priority order that minimizes bandwidth usage.

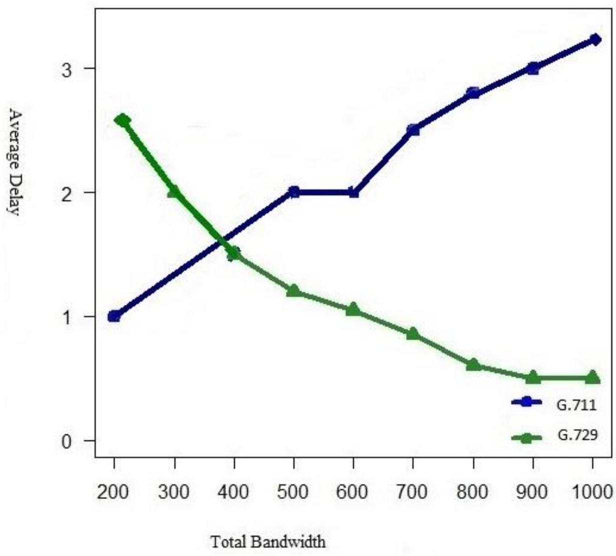

Figure 4. Average delay

ISBN: 978-988-14049-4-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

E. Packet Loss

G.729 codec saw an improvement in packet loss, which is defined as the measurement of packet loss during transmission in a network, while in G.711 packet loss increased during communication. As seen in Figure 5, G.729 codec reduced packet loss since it uses less bandwidth and is compressed when compared to G.711 codec. Because the G.729 codec uses less bandwidth than other G.711 codecs thus, congestion is reduced as a result.

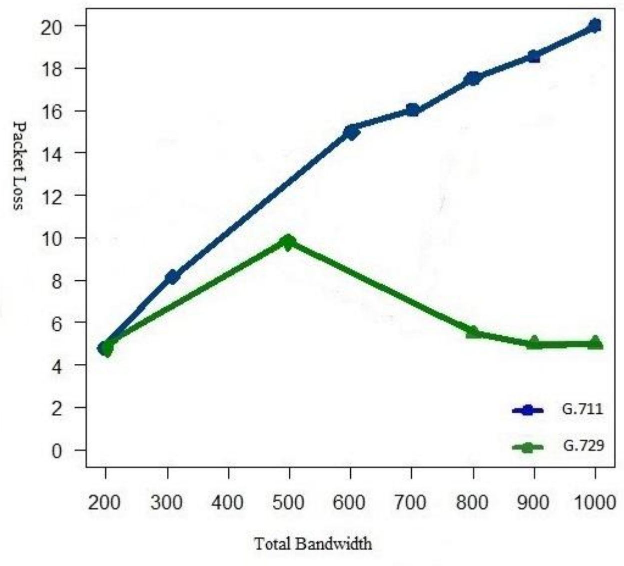

Figure 5. Total packet Loss

VI. Conclusion and Future work

This paper proposed an adaptive codec selection technique that can automatically negotiate codec with other VoIP gateways in order to prevent codec mismatch. Thereby improving the QoS and voice quality. The proposed algorithm was compared with the dynamic codec algorithm which depends on bandwidth and packet loss for adequate data transmission. Experimental research was to evaluate the performance of mostly used VoIP codecs namely G.711 and G.729. The simulation result showed that G.711 is the preferred VoIP codec in terms of VoIP quality however it uses higher bandwidth as compared to G.729.

References

[1] C. Varun, and R. Dudhe, "Transcoding of Voice Codecs G. 711 to G. 729 and Vice-versa Implementation on FPGA." *SSRG International Journal of Electronics and Communication Engineering (SSRG-IJECE)*, Vol. 3, No. 12, pp. 13-17, 2016.

[2] W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135. Choi, Jin-Kyu, et al. "Improvement issues on transcoding algorithms: for the flexible usage to the various pairs of speech codec." *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1. IEEE, 2004.

[3] M. Mikulec, M. Fajkus, M. Voznak, and O. Johansson, "Multiple Transcoding Impact on Speech Quality in Ideal Network Conditions". *Advances in electrical and electronic engineering*, Vol. 13, No. (5), pp. 552-557, 2015.

[4] T.G. Kang, H.J. Bae, D.Y. Kim, and D.U. Kim, "SIP/SDP signaling of media gateway with transcoding function in converged network". *In The 6th International Conference on Advanced Communication Technology*, Vol. 2, pp. 842-845, February 2004.

[5] M. Naeem, V. Namboodiri, and R. Pendse, "Energy implication of various VoIP codecs in portable devices". *In IEEE Local Computer Network Conference*, pp. 196-199, October 2010.

[6] M.O. Ortega, G.C. Altamirano, and M.F. Abad, "Evaluation of the voice quality and QoS in real calls using different voice over IP codecs". *In 2018 IEEE Colombian Conference on Communications and Computing (COLCOM)*, pp. 1-6, May 2018.

[7] G. Afrizal, "Impact of random and burst packet loss on voice codec g. 711, g. 722, g. 729, amr-nb, amr-wb". I*n 2018 4th International Conference on Wireless and Telematics (ICWT)*, pp. 1-4, July 2018.

[8] S.C. Satapathy, V. Bhateja, and S. Das, "Smart intelligent computing and applications". *In Proceedings of the second international conference on SCI*, Vol. 1, pp. 1 – 7, 2018.

[9] S Kirillov, and V. Dmitriew, "A complex algorithm frro objective evaluation of the decodec speech signal quality under the action of acoustic interference". *Journal of Communications*, pp. 34-55.

[10] T.E. Mathonsi and O.P. Kogeda, "Handoff delay reduction model for heterogeneous wireless networks". *In 2016 IST-Africa Week Conference*, pp. 1-7, May 2016.

[11] N. Bayer, B. Xu, V. Rakocevic, and J. Habermann, "Application-aware scheduling for VoIP in wireless mesh networks". *Computer Networks*, Vol. 54, No. 2, pp. 257-277.

[12] T.E Mathonsi, O.P Kogeda and T.O Olwal, "Intersystem handover decision model for heterogeneous wireless networks". *In 2018 Open Innovations Conference (OI)*, pp. 1-7, October 2018.